Assessment of insurance coverage and claims in rainfall related risks in processing tomato in Western Spain

A. Castañeda-Vera, L. Barrios, A. Garrido, I. Mínguez

A B S T R A C T

An extension of guarantees related to rainfall-related risks in the insurance of processing tomato crops has been accompanied with a large increase in claims in Western Spain, suggesting that damages may have been underestimated in previous years. A database was built by linking agricultural insurance records, meteorological data from local weather stations, and topographic data. The risk of rainfall-related damages in processing tomato in the Extremenian Guadiana river basin (W Spain) was studied using a logistic model. Risks during the growth of the crop and at harvesting were modelled separately. First, the risk related to rainfall was modelled as a function of meteorological, terrain and management variables. The resulting models were used to identify the variables responsible for rainfall-related damages, with a view to assess the potential impact of extending insurance coverage, and to develop an index to express the suitability of the cropping system for insurance. The analyses reveal that damages at different stages of crop development correspond to different hazards. The geographic dependence of the risk influences the scale at which the model might have validity, which together with the year dependency, the possibility of implementing index based insurances is questioned.

Keywords:
Processing tomato
Risk
Rainfall
Crop insurance
Index insurance
Extremadura
Spain

1. Introduction

Agricultural insurance systems are plagued by numerous problems that threaten their economic performance and continuity. Poorly designed policies, the effect of asymmetric information (where insurance purchasers have more information than the insurance companies about actual risk and behaviour), unfair loss adjustment procedures, and biases in setting premia are among the main factors that have caused agricultural insurance failures (Binswanger-Mkhize, 2012; Mahul and Stutley, 2010; Gulseven et al., 2011). As in most branches of insurance, the loss or damage of the insured good must be a direct effect of an unambiguously measurable event, which in the case of cropping is the partial or complete loss of harvest caused by an observable climatic hazard or anomaly.

Since the passing of the agricultural insurance law in 1978 (Antón and Kimura, 2011), the aim in the Spanish insurance policy has been to provide a broad and extensive coverage to all agricultural and livestock production enterprises (Antón and Kimura, 2011). The practical objective is to stabilize income in the primary sector by providing protection against adverse climatic events. Risks covered for plant production are mainly related to climatic adversities. Most important claims between 2003 and 2008 were related to hailstorm events with 43% of the reported claims, followed by frosts (19.3%), droughts (14.4%), wind (8.4%) and rainfall (7.4%) (Ruiz-Zorrilla, 2010).

Processing tomato farmers have benefited from expanded coverage since it was first offered in 1990. Additional risks related to rainfall have been progressively incorporated and the coverage has been extended to longer periods during crop development. It now includes damages caused by frost, hailstorms, persistent rainfall through the entire crop cycle and also floods caused by torrential rainfall. The insurance guarantees the production till mid-October and for a maximum of five months of the crop cycle. Damages from hailstorms are considered separately due to their distinct characteristics in frequency, scope, and type of damage. These expansions of coverage have been accompanied by a significant increase in claims in recent years. This suggests that damages related to heavy (over 40 mm in 24 h) or persistent rainfall (over a week with consecutive rainy days each with more than 0.2 mm) may have been underestimated. However, since the damages

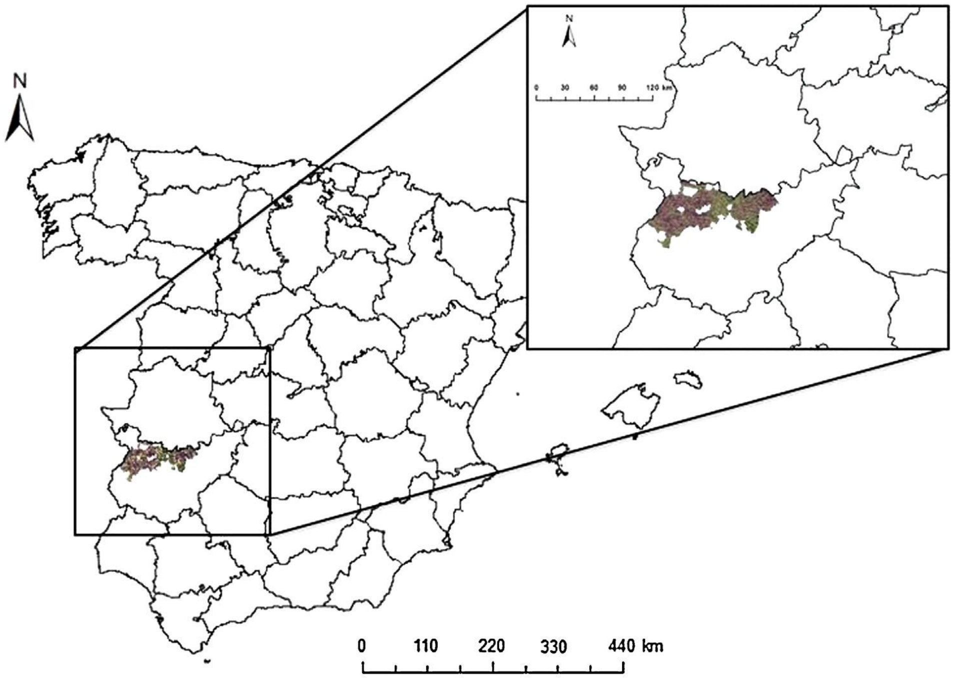

Fig. 1. Study region framed in the middle Guadiana river basin (W Spain).

covered by the current processing tomato insurance are insufficiently defined quantitatively in terms of meteorological variables, it is difficult to assess the frequency and magnitude of the crop losses.

If rainfall risk could be defined in terms of measurable weather variables, indemnities might be estimated by applying mathematical formulae to the observed weather data recorded in a given weather station, to which the local farmers purchasing the insurance policy would be associated. This could have two advantages. First, it might facilitate the valuation of current and future risk by using long weather data series and future climate projections. Second, it might simplify the payment of indemnities by improving the current system of visual (and/or sampling) assessment by experts, as it is currently done. Claims under such a Weather Index Insurance (WII) would be indemnified based on values obtained from an index that serves as a proxy for losses. Success of such a system requires that the index for observed weather events is strongly correlated with crop losses on all farms; otherwise it would produce false positives and false negatives. This is one of the reasons why WII has been little adopted in all countries where it has been offered commercially (Binswanger-Mkhize, 2012).

The objective of this work was to define the incidence of rainfall-related risks for processing tomato production in the Extremenian Guadiana River basin of W Spain in terms of measurable and objective variables. First, the risk related to rainfall was modelled as a function of meteorological, terrain and management variables using data from the recent period of broadest insurance coverage. Second, the resulting model was applied to the previous period with more restricted coverage in order to evaluate the possible impact of the extended coverage on the observed increase in claims. Results are used to discuss the opportunities to insure processing tomato against rainfall risks through a WII. To our knowledge, this is the first attempt to check the potential of WII for horticultural crops.

2. Materials and methods

2.1. Current damage appraisal for insurance payments

Indemnities are based on in-field evaluation made by loss adjusters of Agroseguro, the pool of insurance companies, following explicit monitoring procedures (Antón and Kimura, 2011). The procedure has the following steps: (a) the farmer files a notification of the event considered to have caused the observed crop loss; (b) Agroseguro checks the occurrence of the event and arranges the visit of a loss adjuster to the farm; (c) the adjuster gathers data, photos, and takes samples, recording all data in a tablet computer; (d) the adjuster either calculates the indemnity while at the farm or sends an offer within a few days; (e) the farmer can accept or reject the offered indemnity; (f-1) if it is accepted the payment is transferred within two months; (f-2) if it is rejected, Agroseguro reviews the adjustment and sends a new indemnity proposal; (g) if the farmer does not accept it, he/she must file a lawsuit. This whole process is subject to quality control procedures independently performed by both Agroseguro and the Consorcio de Compensación de Seguros, which is the public reinsurance entity in Spain.

2.2. Processing tomato sector

Spain is the fourth largest producer of processing tomato in the world, behind USA (California), China, and Italy. About 70% of the national production is harvested in the Guadiana River Basin in Extremadura, Western Spain (Macua-González et al., 2012). Yields average 80 $\mathrm{Mg\,ha^{-1}}$ fresh weight while best fields and farmers achieve 100 $\mathrm{Mg\,ha^{-1}}$.

The study region is located in the middle of the production area (Fig. 1) in the province of Badajoz, W Spain. Summers are warm and relatively dry while winters are mild and wet. The average monthly rainfall and mean maximum and minimum temperatures for the town of Badajoz obtained from AEMET (the Spanish Agency for Meteorology) are presented in Fig. 2. The average altitude in the western area is 285 m and the dominant soils have silty clay texture (Badajoz County). The eastern area has an altitude of 430 m and the soils have a gradually greater content of sand (Don Benito County).

In the study area, production is organized through a technical committee that comprises tomato farmers, cooperatives and about fifteen tomato industries of different size that process the whole production of the region. This committee negotiates total cultivated area, prices, cultivars, quality standards and delivery calendars.

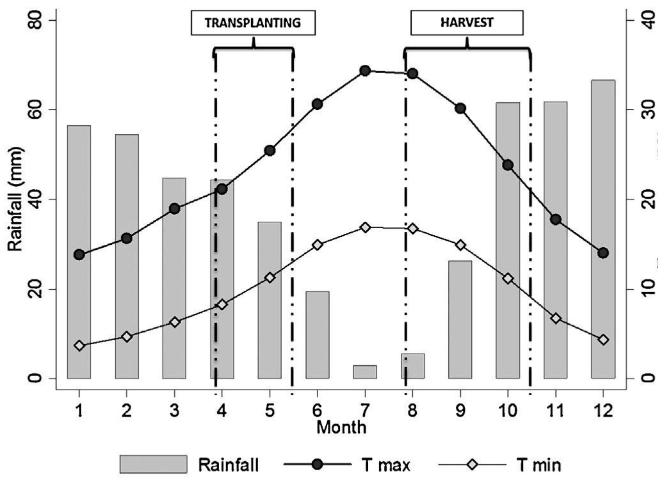

Fig. 2. Climodiagram corresponding to Badajoz. Average values for the period 1955–2011.

These decisions are negotiated in order to facilitate interactions between farmers and industry, and the organization of the harvest season. Vertical coordination in the sector ensures that eligibility criterions to receive the subsidies from the Common Agricultural Policy are met. Homogeneity of production is based on a selection from the long list of available varieties as proposed by the technical committee. The cooperatives organize the planting and delivery calendars in order to share the risks of an early or late planting among all farmers holding production quotas. Other decisions, including irrigation and fertilization, are taken by individual farmers. Input prices and adverse climate events remain uncontrollable.

Processing tomato is grown between April and October. Planting dates are scheduled and assigned to farmers to distribute the harvest over the two-month period when the processing industry is operating. Planting is also conditioned by the weather conditions because earliest planting dates are limited by frosts and latest by autumn rainfall. Harvest begins at the end of July and finishes at latest by early October.

Main weather hazards for field-grown tomato are hail and extreme rainfall. Tomato is very sensitive to high soil moisture content. Extreme rainfall events may cause flooding in heavy soils or flat areas. Although drip irrigation is gaining importance, furrow or flood irrigation remains common because land is levelled for flooded rice that is included in the crop sequences. Damages associated with waterlogging are hypoxia that inhibits respiration of roots and causes accumulation of endogenous ethylene. With rising temperatures, the effects of flooding lead to wilting and death of plants. The accumulation of surface water during warm days increases relative humidity within the crop encouraging the growth and infection by bacterial and fungal diseases, such as powdery mildews (Guzman-Plazola et al., 2003).

Proper drainage and levelling of fields and use of resistant varieties are among the management practices proposed to avoid problems related to waterlogging. Adverse weather exacerbates disease problems because phytosanitary treatments are not applied unless there is evidence of disease or unless rainfall is forecast during sensitive crop growth periods.

Risks related to rainfall have been covered by insurance since 1999. Initially, damages were indemnified only for crops that had reached fruit setting. The guarantee period was extended in 2010 to include damage occurring beforehand. In present insurance contracts, rainfall-related hazards are catalogued as “torrential rainfall”, “persistent rainfall”, or just “rainfall”, with no objective definition in terms of quantity or frequency. Crop loss is related to fruit wilting or physical damage to vegetative tissue hampering crop recovery. Indemnities are only paid when damages affect more than 20% of the insured tomato yield, which is based on local expected yield considering the insurance records.

Risk is the product of hazard and vulnerability, whereas claim is the record of an incident after a farmer reports crop damage due to a weather event. In this paper, the risk during the growth of the crop is treated differently to that at harvesting. Only records that resulted in an indemnity payment are included as claims in this study.

2.3. *Insurance database*

Data of all processing tomato farmers insured in the region from 1990 to 2011 were made available by the agricultural insurance company (*Agroseguro*). Individual insurance records include information of field location, insurance contract date, insured yield and field size, cultivars used, risks suffered and date. Several filters were applied to the database in order to avoid duplicate records and typing mistakes. Duplicates that were identified by comparing farmer’s and field’s IDs for individual years were deleted. Lastly, fields insured before 2006 were also deleted because they did not contain transplanting date. *Agroseguro* protocols for ensuring data confidentiality were followed.

Fig. 3 shows the frequency of claims after transplanting and the degree day summation by the crops (°C d, base temperature of 10 °C). Two periods can be differentiated, one between days 30 and 60 after transplanting, corresponding to 300–600 °C d, and the second between days 100 and 130 after transplanting, corresponding to 1200 °C d. Risks were identified according to these individual periods. Claims before day 90 after transplanting were coded as related to risks during crop growth (C_Growth), while claims registered in crops that had reached 90 days after transplanting were coded as related to risks during harvesting (C_Harvest).

The three municipalities, Badajoz, in the West, Mérida and Don Benito, in the East, were selected as the largest producers from 2006 and 2011. The corresponding weather stations were Bercial, Mérida, and Palazuelo, for which data are available in the System of Agroclimatic Information for Irrigated crops (SIAR). A database of 16353 observations was constructed with each entry corresponding to an insured field that were not, however, homogenously distributed between the three counties. About two thirds of the data corresponded to fields located in Don Benito, followed by *ca.* one third in Badajoz, and lastly, Mérida, with the smallest number. The incidence of rainfall-related damages differed between years and county ranging from zero to 23.3% of the insured fields registering a claim (Table 1).

The regression model includes continuous and categorical variables that represent different components of the system, *viz.* management, weather, topography. The management variable (TrD) to include the discussion on adaptation strategies; weather variables that explained risk occurrence would constitute the basis to design a WII; the significance of topographic variables would suggest the need to segment insurance conditions depending on the physical location of the fields; the County variable was included to evaluate the scale at which the model was valid; lastly, significance of the Year variable was oriented to assess the time-specificity of the model and, therefore, test the limitation of the cropping system to be insured by a WII.

Weather and topographic variables were included by cross-checking field data with meteorological data from local weather stations and a digital elevation model. In addition new variables were calculated for slope and accumulation of runoff. These variables were calculated with the corresponding tool in the ESRI ArcMap 10.0 GIS over the Digital Elevation Model facilitated by the Service of Digital Cartography and Spatial Data

Table 1
Fields insured and percentage of claims during growth and at harvest per year and county.

Year	County								
	Badajoz			Mérida			Don Benito		
	n	C_Growth	C_Harvest	*n*	C_Growth	C_Harvest	*n*	C_Growth	C_Harvest
2006	580	0.00	0.00	76	3.95	0.00	790	0.38	0.00
2007	451	0.67	15.44	97	0.00	13.40	1405	0.07	1.94
2008	754	1.19	1.85	179	1.68	4.90	1335	3.52	0.00
2009	1275	0.00	0.00	216	0.00	0.00	2136	0.00	0.00
2010	1562	0.55	1.14	253	1.85	0.00	2623	1.64	0.86
2011	759	11.91	8.30	179	23.32	0.00	1683	2.74	0.21
Total	5381			1000			9972		

Source: Agroseguro.

Table 2
Description of the dependent and independent variables included in the initial analysis.

Variable	Description
C_Growth	Claims during the growth of the crop
C_Harvest	Claims at harvesting
Pr1	Maximum daily precipitation between days (30 and 60) – (100 and 130) after transplanting
Pr2	Accumulated precipitation between days (30 and 60) – (100 and 130) after transplanting
Pr3	Variance in daily precipitation between days (30 and 60) – (100 and 130) after transplanting
Pr4	Maximum of consecutive days raining between days (30 and 60) – (100 and 130) after transplanting
Pr5	Maximum of the accumulated precipitation in five consecutive days between days (30 and 60) – (100 and 130) after transplanting
Pr6	Maximum daily precipitation in the crop growth period
Pr7	Accumulated precipitation in the crop growth period
Te1	Maximum of the maximum daily temperature reached between days (30 and 60) – (100 and 130) after transplanting
Te2	Minimum of the minimum daily temperature reached between days (30 and 60) – (100 and 130) after transplanting
Te3	Variance of the maximum daily temperature reached between days (30 and 60) – (100 and 130) after transplanting
Te4	Variance of the minimum daily temperature reached between days (30 and 60) – (100 and 130) after transplanting
Te5	Maximum of the moving average of the maximum temperature (5 d) between days (30 and 60) – (100 and 130) after transplanting
Te6	Minimum of the moving average of the minimum temperature (5 d) between days (30 and 60) – (100 and 130) after transplanting
TrD	Transplanting date (Day of the year)
To1	Average slope in the field
To2	Minimum slope in the field
To3	Average flow accumulation in the field
To4	Maximum flow accumulation in the field
A	Pr4 × Te5
B	Pr5 × Te5
C	Pr5/Pr4
D	Pr4/Te3
E	Pr5/Te3
County	–
Year	–

Table 3
Descriptive statistics of continuous independent variables.

Growth						Harvesting					
Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
TrD	16353	114.1	14.0	91	151	TrD	14216	116.3	14.8	91.0	151.0
Pr1	16353	12.8	9.0	0.2	32.2	Pr1	14216	5.5	6.9	0.2	66.4
Pr2	16353	27.7	19.8	0.2	80.8	Pr2	14216	8.4	10.9	0.2	104.2
Pr3	16353	0.5	0.2	0.2	1.0	Pr3	14216	0.8	0.2	0.2	1.0
Pr4	16353	3.5	2.1	1.0	10.0	Pr4	14216	1.8	1.1	1.0	7.0
Pr5	16353	23.1	14.5	0.2	48.4	Pr5	14216	7.5	9.4	0.2	93.0
Pr6	16353	4.6	2.9	0.0	9.7	Pr6	14216	17.7	7.8	6.0	66.4
–	–	–	–	–	–	Pr7	14216	75.3	38.5	18.4	205.4
Te1	16353	35.9	2.4	23.4	41.8	Te1	14216	38.3	1.7	27.9	42.1
Te2	16353	9.4	2.1	1.7	15.8	Te2	14216	11.6	2.7	0.0	17.6
Te3	16353	4.1	0.8	0.1	6.2	Te3	14216	3.2	0.8	1.1	5.7
Te4	16353	2.5	0.6	0.2	4.4	Te4	14216	2.2	0.5	0.5	4.1
Te5	16353	34.3	2.5	23.0	40.3	Te5	14216	34.7	2.1	28.0	40.3
Te6	16353	17.6	2.0	11.3	24.4	Te6	14216	17.9	1.7	12.5	24.4
To1	16353	3.0	1.6	0.0	17.9	To1	14216	3.0	1.6	0.0	17.9
To2	16353	0.2	0.7	0.0	9.0	To2	14216	0.3	0.7	0.0	9.0
To3	16353	10.8	12.7	0.0	263.3	To3	14216	10.9	13.0	0.0	263.3
To4	16353	238.8	311.6	0.0	3108	To4	14216	235.8	314.2	0.0	3108
A	16353	116.4	65.7	27.3	361.2	A	14216	62.7	38.0	28.0	273
B	16353	775.9	482	7.2	1696.9	B	14216	266.1	334.1	5.6	3470.8
C	16353	7.1	4.9	0.2	40.0	C	14216	4.2	5.0	0.2	34.6
D	16353	0.9	0.6	0.2	7.1	D	14216	0.6	0.3	0.2	2.2
E	16353	5.8	3.8	0.1	38.6	E	14216	2.2	2.6	0.1	30.3

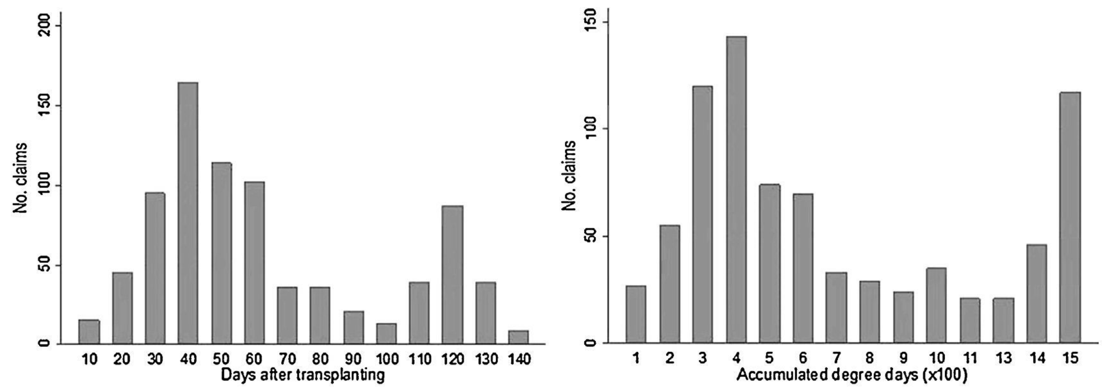

Fig. 3. Claims depending on the days after transplanting and on the accumulated degree days of the crop.

Infrastructure belonging to the University of Extremadura. The variables are described in Table 2. Table 3 provides a description of the continuous independent variables.

Risk incidence varies between years, in part due to changes in the insurance coverage from 2010, so the risk model was fitted to three different datasets of both C_Growth and C_Harvest. The first dataset (*All*) included the insured fields in the three selected counties from 2006 to 2011; the second (*High*) included only the fields insured in the years for which indemnified fields exceeded 1% of the insured fields in at least one of the counties. In this set, greater losses were recorded during crop growth in 2008, 2010 and 2011 but at harvest in 2007 and 2011. The third dataset (*Extended*) included data from 2010 onwards, and therefore only fields insured with the extended risk coverages.

2.4. Risk model

A binary logistic regression model (also known as logit model) was used to investigate the influence of management, topography and meteorological variables on the occurrence of rainfall-related damages for individual fields. The model assumes a binomial distribution for the binary dependent variable and estimates the probability of a claim based on the independent variables of continuous or categorical nature. Indemnified fields are coded "1" and the remainder "0".

The risk model is defined as:

$$log\left[\frac{P(Y=1)}{1-P(Y=1)}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (1)$$

where $P(Y=1)$ and $1-P(Y=1)$ are the probabilities "indemnity" or "no indemnity". It can be rewritten to predict the probability, between 0 and 1, of a claim for individual fields based on the independent variables (x_is) as

$$P(Y=1) = \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}} \quad (2)$$

where the parameter α is the intercept in the model, and β_i are the regression coefficients for each covariate x_i.

The model was fitted with Stata v12 (StataCorp, 2011).

The inclusion of the independent variables in the model was performed using the *stepwise selection procedure*. New variables are added one at a time (forward steps). After each forward step, the log likelihood values of the included variables are reexamined and any non-significant variables are removed. This process continues until no variables are either added or removed from the model.

Jackknife post estimation (Efron, 1981) was performed in order to ensure that parameter estimation was robust. This procedure improves robustness generating alternative design-based standard errors for hypothesis testing and confidence intervals which reduce reliance on theoretical assumptions. The odd ratio was used to rate the independent variables that better explain the dependent one. This ratio is interpreted as the effect of one unit of change in a variable when the other variables remain constant.

The indices of sensitivity and specificity were used to evaluate the accuracy of the models' fit. They are intrinsic qualities of a test and describe how well the models discern between true positives and true negatives. The sensitivity measures the power of the model to identify positives and the specificity measures the power to identify negatives:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (4)$$

In this case, TP (true positives) is the number of "claims" correctly identified by the model and TN (true negatives) is the number of "no-claims" cases correctly identified by the model. FN (false negatives) is the number of "no-claims" identified as "claims" and FP (false positives) is the number of "claims" identified as "no-claims".

Logit models output a probability of success. This probability is a continuous variable that is transformed into the binary variable *success-no success* using a cut-off value. The cut-off value is the probability at which an observation is to be classified either as positive or as negative. It influences the resulting number of true positives, true negatives, false positives, and false negatives after applying the model to a specific set of data. For a change in the cut-off selected, an increase in the sensitivity will result in a decrease in specificity and *vice versa*. The cut-off selection must reflect the best tradeoff between both indices, and therefore, obtain the optimum allocation of estimated positives and negatives to observed data. As a default, logit models often use a cut-off value of 0.5; for this analysis, we selected the cut-off values 0.2 and 0.8 to evaluate the significance of this parameter in the final classification of expected claims.

3. Results

3.1. Statistical analyses

Models were fitted for each of the damages (C_Growth and C_Harvest) for the three datasets (Tables 4–6). The categorical variables, County and Year, were included to evaluate their marginal contribution to the probability of claims and hence of the potential for development of WII that could only be based on weather

Table 4
Logit models for claims during the growth of the crop and at harvesting.

DS[a]		Variables (*Significant variables in bold, $p \leq 0.05$*)
Growth		
All	2006–2011	**Te1 Pr3 Te6 TrD Te2** ***D*** **To3 Pr2 B Pr5 Te3 Te5 Pr1** *C* **E** *Te4* **To1 Pr4 A**
High	2008|2010|2011	**Te1 Pr3 Te6 TrD Te2 Pr2 D To3 Te5** *C* **Pr1** *Te3* **B Pr5** *Te4* **E A Pr4**
Extended	2010–2011	*Te1* **Pr3** *D* **Te2 TrD E** *Te3* **To3 Pr1 Pr2 C A Pr4 Te5 Pr5** *B*
Common variables		***(Pr1 Pr2 Pr3 Pr4 Pr5) (Te1 Te2 Te3 Te5) (TrD) (A B D E)***
Harvest		
All	2006–2011	**Pr1 Te6 TrD Te2 Te4 Pr7 C Te5 Pr2 D A Pr6 Pr5 Pr4 Te1 To2 To1 To3 Pr3**
High	2007|2011	*C* **TrD Te4 Te2 Te6** *Pr2* **Te5** *Te3* **A** *B* **Pr4 Pr3** *Pr1* **To2 D E** *To3* **To1 Pr7** *Pr6*
Extended	2010–2011	*Pr1* **Pr4 D A Te2 Te4 TrD Pr6 Te6** *Te3* **Pr3** *B* **Pr7 E Te5 Pr5** *Pr2*
Common variables		***(Pr1 Pr2 Pr6 Pr7) (Te1 Te2 Te4 Te5) (TrD) (B)***

[a] DS: Datasets *All* (2006–2011); *High* (Years with higher proportion of damages); *Extended* (2010–2011).

Table 5
Goodness of fit (R^2 statistic) with and without categorical variables included.

DS[a]	Model	n	Categorical variables effect on R^2			
			None	County	Year	County & Year
Growth						
All	2006–2011	16,353	0.56	0.58	0.68	0.69
High	2008|2010|2011	9327	0.66	0.67	0.71	0.71
Extended	2010–2011	7059	0.70	0.71	0.71	0.72
Harvest						
All	2006–2011	15,759	0.38	0.39	0.50	0.50
High	2007|2011	4355	0.48	0.50	0.52	0.53
Extended	2010–2011	6605	0.50	0.51	0.60	0.61

[a] DS: Datasets *All* (2006–2011); *High* (Years with higher proportion of damages); *Extended* (2010–2011).

Table 6
Sensitivity and specificity of the models the categorical variable County included.

DS[a]	Claims (%)/cut-off 0.2		Claims (%)/cut-off 0.8	
	Sensitivity	Specificity	Sensitivity	Specificity
Growth				
All	65.97	99.25	39.81	99.98
High	75.12	98.21	54.50	100.00
Extended	77.41	98.30	64.46	100.00
Harvest				
All	45.07	99.46	10.77	99.98
High	70.80	98.18	28.32	100.00
Extended	71.97	99.46	30.57	100.00

[a] DS: Datasets *All* (2006–2011); *High* (Years with higher proportion of damages); *Extended* (2010–2011)

variables. Fields located in Mérida in 2010 were not included for risks at harvesting as no claims were registered that year in that county.

Table 4 lists the variables included in the models according to their contribution to precision, the significant ones in bold. The common variables for both definitive models for each risk (C_Growth and C_Harvest), present after applying the stepwise selection procedure, are in italics. The weather variables related to variances in daily precipitation (Pr3) and temperatures (Te1, Te2, and Te6) were the first included in the stepwise selection procedure but transplanting date (TrD) was also significant. Topographic variables made a minor contribution to explain risk incidence. Models for dataset *Extended* included fewer variables than for databases *All* and *High*.

Table 5 records the number of fields used to fit each of the models (n) and the goodness of fit (R^2) with and without the categorical variables, County and Year. Their inclusion substantially improves risk analysis of both C_Growth and C_Harvest models. Model *All* shows the largest improvement, with an increase of R^2 from 0.56 to 0.69 for risks during C_Growth, and from 0.38 to 0.50 for C_Harvest. In both cases, the largest effect is attributed to Year.

A classification of the sensitivity and specificity of the models is presented in Table 6. The fraction of correctly classified scenarios was greater when using a cut-off criterion of $p = 0.2$ compared to a cut-off of $p = 0.8$. In the models based on dataset *Extended* (years 2010 and 2011), a cut-off of $p = 0.2$ resulted in a sensitivity of 80% and a specificity over 98%.

Performance of the model for *Extended* dataset for claims during C_Growth with inclusion of the categorical variable County is presented in Table 7. All selected variables for rainfall were significant. For variables with greater odd ratios, risk increases coincide with increases in variance of daily precipitation (Pr3) and increases in the maximum of accumulated precipitation within five consecutive days (Pr5). This suggests that claims are related to periods of intense and concentrated rainfall. The probability of damage diminished as transplanting date was delayed (management). The categorical variable County reveals that compared to the reference

Table 7
Parameters estimate for the logit model for dataset *Extended* explaining claims during the crop growth including the categorical variable County.

Variable	Coefficient	Standard error	p value[b]	Odd ratio	95% CI[a]
Intercept	127.439	28.9	***	–	101.9/198.4
Meteo					
Pr1	−0.5	0.1	***	0.6	−0.7/−0.3
Pr2	0.1	0.0	***	1.1	0.1/0.2
Pr3	34.4	4.2	***	**>1000**	26.1/42.7
Pr4	−23.7	3.9	***	0.0	−31.4/−16.1
Pr5	2.8	0.8	***	**16.0**	1.1/4.4
Te1	−1.4	0.8	*	0.2	−3.0/0.2
Te2	1.4	0.1	***	4.2	1.2/1.7
Te3	0.1	0.6	–	1.1	−1/1.3
Te5	−3.4	1.2	***	0.0	−5.7/−1.0
A	0.7	0.1	***	2.0	0.5/0.9
B	−0.1	0.0	***	0.9	−0.1/0.0
D	−1.2	1.9	–	0.3	−5.0/2.6
E	−1.1	0.3	***	0.3	−1.7/−0.4
Management					
TrD	−0.239	0.0	***	0.00	−0.3/−0.2
County					
Badajoz	Reference	//	//	//	//
Mérida	4.790	1.7	***	**34.1**	0.5/6.5
Don Benito	5.323	1.7	***	**180.2**	2.0/8.4

[a] Coefficients confidence interval.
[b] – $p \geq 0.1$; *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 8
Parameters estimate for the logit model for dataset *Extended* explaining claims at harvesting including the categorical variable County (Mérida not included).

Variable	Coefficient	Standard error	p value[b]	Odd ratio	95% CI[a]
Intercept	−50.3	−5.3	***	–	−68.8/−31.8
Meteo					
Pr1	0.5	8.0	***	1.6	0.4/0.6
Pr2	−0.1	−3.0	***	0.9	−0.2/0.0
Pr5	−2.5	−7.5	***	0.1	−3.2/−1.9
Pr6	0.2	1.3	–	1.2	−0.1/0.4
Te1	−0.2	−1.5	–	0.8	−0.4/0.1
Te2	1.5	10.5	***	**4.5**	1.2/1.8
Te4	4.8	9.0	***	**116.2**	3.7/5.8
Te5	−0.1	−0.4	–	0.9	−0.3/0.2
B	0.0	2.2	***	1.0	0.0/0.0
Management					
TrD	0.2	4.9	***	1.2	0.1/0.3
County					
Badajoz	Reference	//	//	//	
Don Benito	1.6	3.3	***	5.2	0.7/2.6

[a] Coefficients confidence interval.
[b] – $p \geq 0.1$; *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

category, Badajoz, risk was significantly higher in Mérida and Don Benito.

Performance of model for dataset *Extended* in explaining claims at harvesting and including the categorical variable County is presented in Table 8. Variables with higher odd ratios for daily precipitation (Pr1, Pr6) and minimum temperature (Te2, Te4) suggest that claims coincide with heavy rainfall and high temperatures. In flat fields, prone to waterlogging, this is related to an increase in the relative humidity. Once again the probability of damage was greater when transplanting date was delayed. Coefficients of the categorical variable County that were estimated relative to Badajoz (the reference group) reveal that the number of claims was significantly higher in Don Benito. No conclusion could be drawn in Mérida because the number of claims was too small.

3.2. *Extrapolating risk*

Fig. 4 shows the application of the *Extended* statistical model, including the categorical variable County, to claims during crop growth and at harvesting between 2006 and 2011. Risk at harvesting was not estimated for Mérida since the model could not be validated for this county. Differences between observed and estimated risk differed between counties during this period. Overall, estimated risk during C_Growth was greater than the registered claims during 2006 to 2009. The overestimation was significant in 2008 in Badajoz (1.19% *vs.* 49.2%), in 2006 and 2008 in Mérida (3.95 *vs.* 28.94, and 1.68 *vs.* 35.75%, respectively), and in 2007 in Don Benito (0.07% *vs.* 31.10%). Estimated risk at harvest showed less differences with respect to registered claims, except for 2007 in Badajoz, where risk was underestimated (15.44% *vs.* 0.45%).

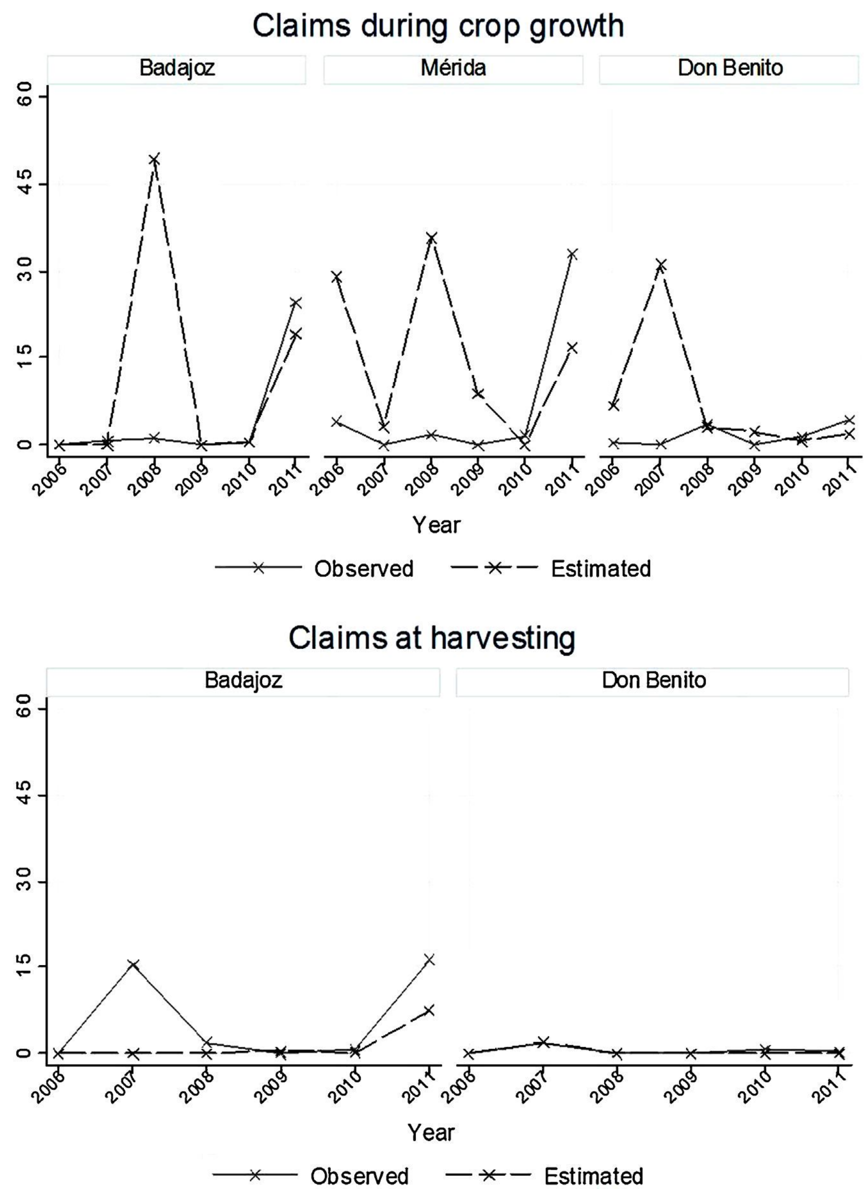

Fig. 4. Estimated and observed field affected by rainfall during the growth of the crop.

4. Discussion and conclusions

The main objective of this paper was to assess the opportunities of protecting farmers of proccessing tomato against rainfall-related risks through a weather insurance index.

Claims registered during the growth of the crop (C_Growth) and those at harvesting (C_Harvest) were modelled separately to account for different crop vulnerabilities and different management decisions, even if in both cases the claims were. The first period corresponds to crop growth and development to flowering and early fruit formation, while the second period coincides with the end of fruit maturation up to harvesting (Onofri et al., 2009). Rao and Li (2003) reviewed the physiological effects of flood duration and growth stage on tomato crops. They established greater sensitivity to flooding during flowering and that damages were related to ethylene accumulation and stomatal closure. At the end of crop development, flooding conditions hasten fruit maturation and decrease quality. Moreover, rainfall at harvesting prevents machinery from operating in fields at the optimal time, producing crop over maturation and consequently yield loss.

The differences in the most significant variables of the models, extreme temperatures and daily precipitation variance, during crop growth and at harvesting suggest that damages at different stages of crop development correspond indeed to different risks. Fields of processing tomato are more vulnerable when exposed to intense and concentrated rainfall during C_Growth, while at harvest they are threatened primarily by rainfall events combined with high temperature. The influence of high temperatures after rainfall events may be related to an increase in relative humidity that favours fungal diseases reducing crop growth and fruit quality (Guzman-Plazola et al., 2003). These conditions also accelerate fruit maturation and shorten the period the period the period that the farmer has before yield is lost. The transplanting date influences risks during both C_Growth and C_Harvest periods. Early transplanted plants are more exposed to damages during vegetative growth, while transplanting at later dates rises the risks at harvesting. Model results suggest, unsurpringly, that shortening the cropping period by avoiding the earliest and latest transplanting dates would decrease risk. However, changes in harvest dates must be coordinated with the industry.

Agricultural risks can be characterized by their covariance, frequency and severity of damages (Hazell et al., 2010). The model improves substantially when categorical variables are included, and the greatest R^2 was obtained with database *All*, for risk both during crop growth and risks at harvesting. This suggests that risk incidence is geographic- and time-specific, a finding that may explain why the Weather Index Insurance (WII) has found limitated application for many crops and why it is little used in most places around the world where it is offered commercially (Binswanger-Mkhize, 2012).

Covariance is related to the degree at which claims are correlated across fields within a region. Meteorological data used in this analysis were obtained from regional stations rather than measured weather in individual fields. In a WII, weather variables operate as derivatives through which a farmer is indemnified depending on the weather at a certain moment rather than from directly measured damages. This means that there must be a trade-off between accuracy and practicality to avoid major discrepancies between actual losses and weather derivative indemnities; this is a common risk when using weather derivatives (Vedenov and Barnett, 2004). The significant influence of the categorical variable County, suggests that the risk under study has a medium to low covariate because the rainfall event impacts heterogeneously in the region. This geographic dependence suggests that extreme rainfall events cause fairly localized damage, influencing the scale at which the model might have validity. This, together with the year dependency, questions the potential of implementing index-based insurance for processing tomato rainfall extreme events.

A logit regression was selected to model the frequency rainfall-related risks in field tomato crops in order to identify the variables that significantly impact the probability of registering a claim. The models estimated in this paper explain the likelihood of a claim being registered for a tomato field as a function of weather, topography, transplanting date, and location. The objective of the model was not to establish cause–effect relationships for each independent variables but to measure the overall probablity of a claim. Weather variables were then built over the periods with the highest probabilities of registering a claim, and not to the previous days of each of the observed claims. Therefore, in this work the models are used to study the concurrence of a weather event and the probability of registering a claim.

Although the coefficient of determination is moderately high (~0.75) in models for database *Extended* when it includes the categorical variable County, specificity approaches 100% and sensitivity exceeds 75% when using a cut-off of 0.2 ($p > 0.8$). That is to say that a non-negligible proportion of observations could not be explained by the model. This might be associated to the low number of claims in proportion to the total number of insured fields and also to possible sparseness and/or multicollinearity of the data. Due to the unbalanced number of observations between counties, it could be expected that the model captures better the reality of the most represented ones.

Agricultural risks related to extreme climatic events have a low frequency and high severity (Binswanger-Mkhize, 2012). Therefore, the small number of successes is an inherent characteristic to the rationale of insurance. From a statistical point of view, insured crop claims might be then treated as "rare events", as recommended by sound actuarial criterion (Seog, 2010). In our case, the range is between 0 and 35%, depending on the year and location. According to King and Zeng (2001), it is difficult to model these cases for two main reasons. First, logistic models tend to underestimate the probability of occurrence of these events. Second, data collection becomes harder when success events are infrequent and so datasets often result a huge number of observations and relatively few and poorly measured explanatory variables.

Claims increased from 2010, in parallel to the extension of insurance coverage. The model fitted for database *Extended* (years 2010 and 2011) was used to evaluate how many claims would have been during 2006 to 2009 if the broadest insurance coverage scenario had been in place. The differences between simulated risk for the broadest coverage insurance and observed damages in 2006 to 2009 indicate that the extended guaranteed period does include actual threats that had been previously afforded by the farmer. From an actuarial point of view, risk assessment must accept that historical data may be insufficient to establish the real risk the insurance is covering. Moreover, the significant increase in indemnities in 2011 are not just related to an unusual year in terms of the precipitation pattern, but are also a consequence of the new integrated guarantees, as initial insurance coverage did not protect farmers when damages occurred before fruit setting.

Lastly, a WII requires an accurate model of the severity of damages (percentage over the insured yield) through objective and measurable variables. Logit model results point at a scale and year dependency of the risk. This suggests that data measured at the regional level are not able to fairly reproduce individual field risks. Modelling the severity of damages requires more detailed variables at field level, in terms of more detailed management information of transplanting date, field history, and especially weather data measured in individual fields.

In this work, insurance data included mainly variables related to location, transplanting date, and expected yield. Meteorological and topographic variables were not available in the insurance database and therefore it had to be crosschecked using transplanting dates and location. The criterion of associating a weather station according to an administrative border obviates the field-to-field variances. This results in low sparseness and high probability of multicollinearity of the meteorological data for fields with similar transplanting dates. This limits the power of the model to explain risk incidence. The model was also limited in the number of field specific variables. There might be soil or management characteristics with a stronger influence on damage occurrence, which are absent in the insurance records datasets. The possibility of implementing Weather Index based Insurances (WII) is hampered by the significant dependency of the model on scale and year. Therefore, our work does not recommend the use of WII to crops that are very sensitive to weather anomalies and crop management.

Acknowledgements

This research was funded by the project MULCLIVAR, from the Spanish Ministerio de Economía y Competitividad (MINECO) CGL2012-38923-C02-02, A. Castañeda-Vera has a PhD grant from UPM. We thank *Agroseguro* and ENESA (Entidad Nacional de Seguros Agrarios) for the data used in the analyses, and Finca La Orden-Valdesequera, Cooperativa Casat (Don Benito, Badajoz) and the Centro Tecnológico Agroalimentario de Extremadura (CTAEX) for the clarifying information regarding processing tomato in Extremadura.

References

Antón, J., Kimura, S., 2011. Risk Management in Agriculture in Spain. OCDE Food, Agriculture and Fisheries. Working paper no. 43.

Binswanger-Mkhize, H.P., 2012. Is there too much hype about index-based agricultural insurance? J. Dev. Stud. 48 (2), 187–200.

Efron, B., 1981. Nonparametric estimates of standard error: the Jackknife, the bootstrap and other methods. Biometrika 68 (3), 589–599.

Gulseven, O., Kucuksenel, S., Nas, S.E., 2011. Trends in agricultural insurance: past, present and future. Actual Problems in Economics 120, 316–320.

Guzman-Plazola, R.A., Davis, R.M., Marois, J.J., 2003. Effects of relative humidity and high temperature on spore germination and development of tomato powdery mildew (*Leveillula taurica*). Crop Protect. 22 (10), 1157–1168.

King, G., Zeng, L., 2001. Logistic regression in rare events data. Political Anal. 9 (2), 137–163.

Macua-González, J.I., Lahoz-García, I., Calvillo-Ruiz, S., Bonzal-Yaguas, J.M., 2012. Variedades de tomate para industria: Resultados de la campaña 2012. INTIA–Cultivos Hortícolas.
Mahul, O., Stutley, C.J., 2010. Government Support to Agricultural Insurance. Challenges and Options for Developing Countries, Washington, DC.
Onofri, A., Becaffi, C., Benincasa, P., Guiducci, M., Tei, F., 2009. Is CropSyst adequate for management-oriented simulation of growth and yield of processing tomato? J. Appl. Hort. 11 (1), 17–22.
Rao, R., Li,F Y., 2003. Management of flooding effects on growth of vegetable and selected field crops. HortTechnology 13 (4), 610–616.
Ruiz-Zorrilla, J., 2010. La experiencia española en el aseguramiento y gestión de riesgos. In: La gestión de riesgos y crisis en el seguro agrario. Ministerio de Medio Ambiente y Medio Rural y Marino, pp. 1–48.
Vedenov, D.V., Barnett, B.J., 2004. Efficiency of weather derivatives as primary crop insurance instruments. J. Agric. Resource Econ. 29 (3), 387–403.